Beeline Software
User Guide

FOR RESEARCH USE ONLY

ILLUMINA PROPRIETARY
Part # 15016340 Rev A
July 2010

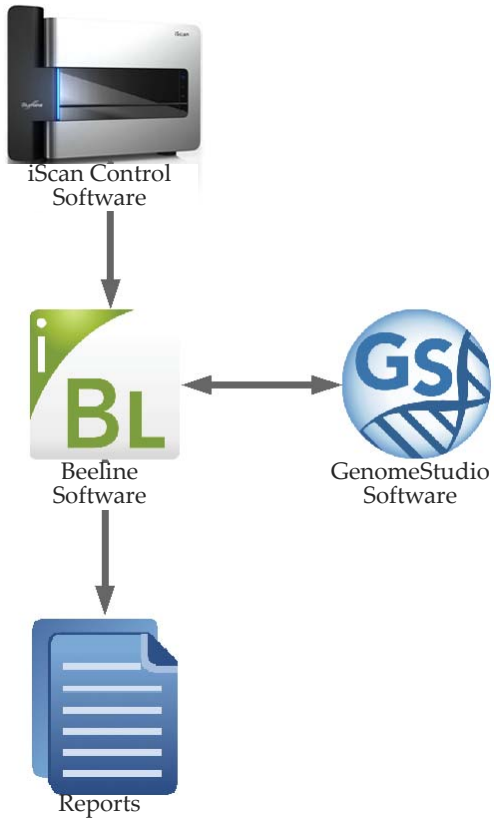
illumina®

Introduction

As the size of experimental data sets increases, both in sample number and chip complexity, the time required to calculate sample statistics and visually interrogate clusters is becoming prohibitive.

Beeline software addresses this bottleneck through the real-time calculation of sample statistics and by reducing experimental data size with flexible filtering capabilities. In short, Beeline offers a direct path to data analysis.

Table 1 Beeline Highlights

Time-Saving Analysis	Enables automatic calculation of sample statistics, allele calling, and cluster file generation
Reduced Data Size	Offers flexible user-defined parameters to select/deselect samples and loci for targeted downstream analysis
Accessible Reporting	Generates text document output of several reports
Seamless Integration	Works directly with Infinium data output and creates projects for direct import into GenomeStudio Data Analysis Software 

Primary Workflow

The diagram below outlines the primary workflow for Illumina's next-generation genotyping analysis. It indicates what files are used as input at each step and what files are produced as output.

NOTE
There are many paths to data analysis. This user guide presents just one example of an integrated end-to-end workflow.

Beeline sits in the middle of this workflow, consuming either raw or normalized data off the instrument and exporting tab-delimited reports and GenomeStudio projects. Beeline is not designed to replace GenomeStudio but rather to augment it, allowing users to reduce total experimental data size and the commensurate analysis time prior to visual analysis in GenomeStudio.

Inputs	Workflow Steps	Outputs	Illumina Software
*.idat	Data Normalization / Sample Statistics	*.gtc	
*.gtc	Filter Samples		
*.bmp *.egt	Filter Loci		
	Create GenomeStudio Subproject	*.bsc	
*.bsc	Cluster Modification (Gentrain)	*.egt	
	Generate Reports	*.csv	

iScan Control Software Beeline Software GenomeStudio Software

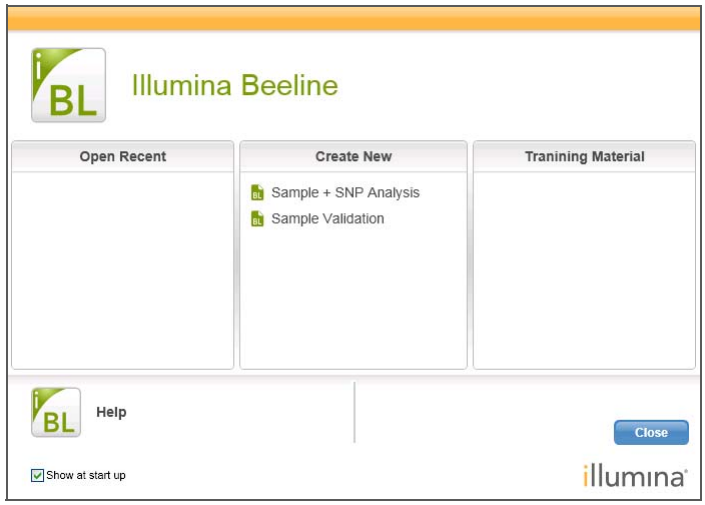
Getting Started

To open the Beeline software:

- Double-click the Beeline icon on your computer desktop, or
- Select **Start** | **All Programs** | **Illumina** | **Illumina Beeline**.

The Beeline Welcome screen opens:

Figure 1 Illumina Beeline Welcome Screen



From the Welcome screen, you can:

- Re-open a recent project by selecting a project from the Open Recent list. The project will open in the main window.
- Start a new project by following the instructions in *Importing Data* on page 6 of this User Guide. The project will open in the main window.
- Review training materials or user documentation.
- Advance directly to the main window without making any selections by clicking **OK**.

TIP

You can configure the software to open directly to the main window by clearing the **Show at start up** check box. If you do so, you can always re-open the Welcome screen by selecting **Help** | **Welcome Screen** from the main window.

Importing Data

Beeline works with normalized data saved in *.gtc files. (Unlike GenomeStudio, Beeline does not read sample statistics from *.idat files.) The data import process described here begins after your *.idat files have been converted to *.gtc files using the new AutoConvert tab in the Options dialog box of the iScan Control Software (ICS) running on the iScan or the AutoCall feature in Illumina LIMS.

NOTE
If you do not have version of ICS nor Illumina LIMS, you can convert *.idat files to *.gtc files using the AutoConvert feature in Beeline before importing data to create your project.

Beeline supports two types of projects:

- **Sample Validation Project**—Useful in a production environment for removing outlying samples from your data before reporting or downstream analysis. Beeline will not import loci statistics for this type of project.
- **Sample + SNP Analysis Project**—Enables you to filter your data on samples and loci and, optionally, modify clusters in GenomeStudio before reporting or downstream analysis.

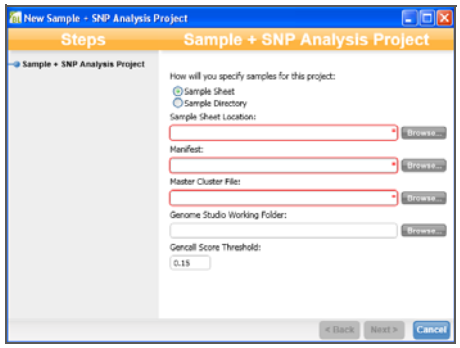
Before you begin, decide which type of project you want to create.

1 From the Create New area of the Welcome screen, click **Sample + SNP Analysis** or **Sample Validation**, as desired.

TIP
If the Welcome screen is not open, you can also start a new project by selecting **File** | **New Project** | **Sample + SNP Analysis** or **Sample Validation**, as appropriate.

The appropriate New Project dialog box opens.

Figure 2 New Sample + SNP Analysis Project Dialog Box



Required fields are highlighted in red and marked with an asterisk (*).

2 Do one of the following:

- If you have a sample sheet already created for your samples, select the **Sample Sheet** radio button, click **Browse** next to the Sample Sheet Location field, and navigate to the sample sheet.

NOTE
Your sample sheet must include a column titled **Path**, identifying, for each sample, the folder location of that sample's *.gtc file.

- Otherwise, select the **Sample Directory** radio button, click **Browse** next to the Sample Directory Location field, and navigate to the folder holding your *.gtc files.

3 If you are creating a Sample + SNP Analysis project, do the following:

a Click **Browse** next to the Manifest field, and navigate to the *.bpm file for this data set.

b Click **Browse** next to the Master Cluster File field, and navigate to the *.egt file for this data set.

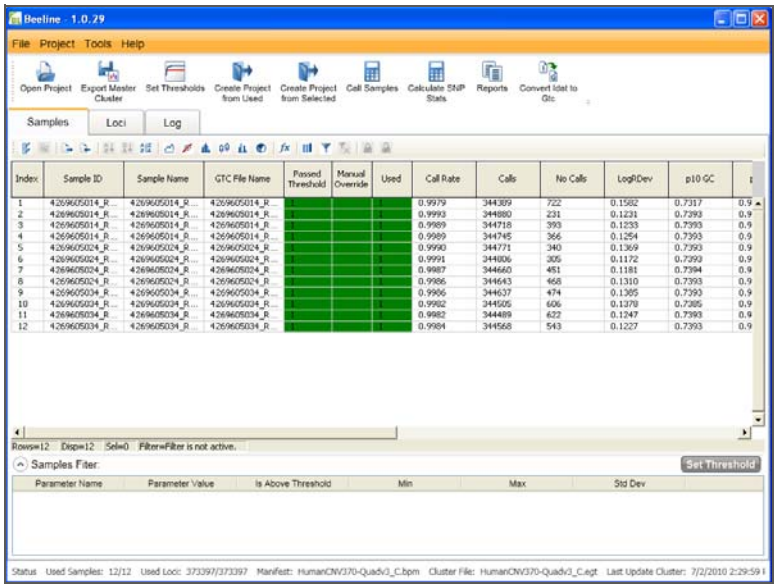
c [Optional] Click **Browse** next to the GenomeStudio Working Folder field and navigate to the folder where you want projects to be saved.

NOTE
If you do not select a working folder now, or if you want to save a project in a different folder later, you can specify the folder when you create the project.

d [Optional] Specify a **Gencall Score Threshold** to use for this project.
Illumina recommends you use a GenCall Score cutoff of 0.15 for Infinium products.

4 Click **Finish**.
Your project data opens in the main window.

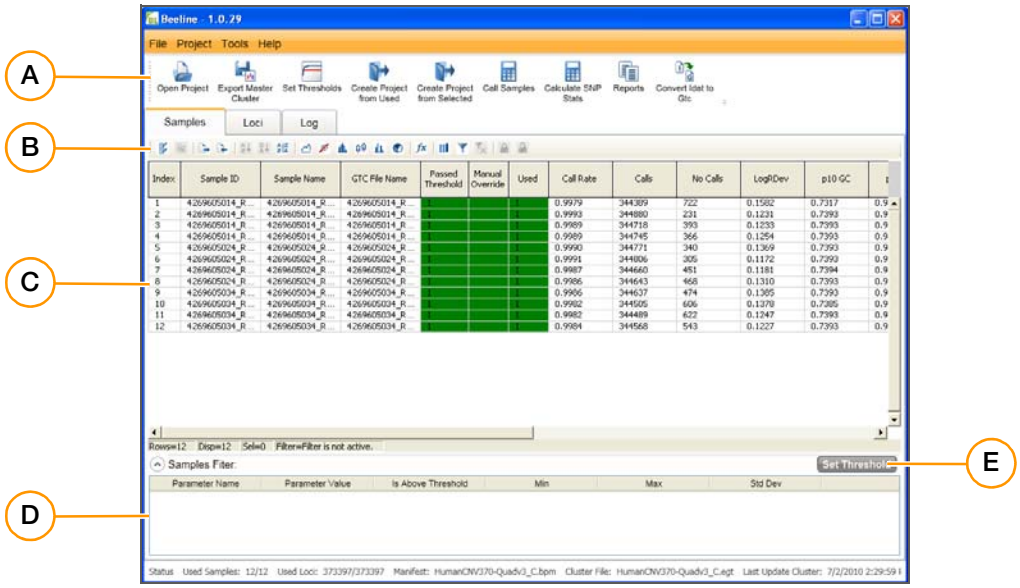
Figure 3 Illumina Beeline Main Window



Overview of the Main Window

The Beeline main window is divided into the following areas:

Figure 4 Overview of the Main Window



A **Workflow Controls**—Enable you to open and filter data, create GenomeStudio projects, and generate reports.

B **Data Display Controls**—Enable you to change the way that your data appears in the statistics tables.

C **Data Tables**—Display your sample and loci statistics. If you selected a Sample Validation project, the data in the Loci Statistics table shows zeroes.

D **Threshold Parameters Window**—Displays the list of thresholds currently set on your data.

E **Set Thresholds Button**—Opens the Sample or Loci Threshold dialog box, where you specify criteria for filtering samples and loci in or out of a project or report.

Data Display Controls

The following buttons are available on the Data Display Controls toolbar for changing the way that your data appears in the statistics tables. These controls do not modify the data in your project.

Feature	Icon	Purpose
Calculate		Updates all statistics in the table. Appears in the data display toolbar only when statistics are out of sync (for example, when data have been filtered out.)
Select all		Highlights all the rows in the data table.
Copy selected rows to the clipboard		Copies selected rows or columns to the clipboard.
Export displayed data to a file		Exports selected rows or columns to a file.

Feature	Icon	Purpose
Import columns into the table		Imports sample or loci data from a file you specify.
Sort column (Ascending)		Sorts the data table in ascending order by the selected column.
Sort column (Descending)		Sorts the data table in descending order by the selected column.
Sort by multiple columns		Enables you to sort the data table by more than one column.
Line plot		Displays a line plot of the sample or loci statistics in the data table.
Scatter plot		Displays a scatter plot of the sample or loci statistics in the data table.
Histogram plot		Displays a histogram of the sample or loci statistics in the data table.
Box plot		Displays a box plot of the sample or loci statistics in the data table.
Frequency plot		Displays a frequency plot of the sample or loci statistics in the data table.
Pie chart		Displays a pie chart of the sample or loci statistics in the data table.
Calculate new column		Allows you to calculate data for creating a new column.
Column chooser		Opens the Column Chooser dialog box for adding or removing columns from the data tables and rearranging the column order.
Filter rows		Opens the Filter Rows dialog box for hiding or showing rows of data based on the criteria you specify. ***This feature does not remove samples or loci from your data set; it affects the way information is displayed in the data table only. If you want to make the data set smaller and easier to work with, you must use the thresholding or manual filtering features described beginning on page 10.***
Clear filter		Removes any filter set with the Filter rows feature so that all rows will appear in the data table. ***This feature does not undo threshold criteria set by following the instructions beginning on page 10; it affects the way information is displayed in the data table only.***
Lock selected columns		"Freezes" a selected column on the left side of the table, so that it remains visible while scrolling through other columns.
Unlock selected columns		"Unfreezes" a selected column so that it scrolls with other columns.

Filtering Your Data

You can filter your project data based on samples, loci, or both to reduce the size of your data set for generating reports or creating GenomeStudio projects.

Filtering Samples

Beeline offers two methods to filter samples. You can set thresholds to automate the filtering process or you can manually include or exclude specific samples.

Setting Thresholds to Filter Samples Automatically

The instructions below provide an example of how to filter samples based on two threshold criteria: p10 GC and call rate. However, you can apply these instructions to filter on any criteria appropriate to your experiment.

1 With the Available Samples tab active, click **Set Thresholds**.

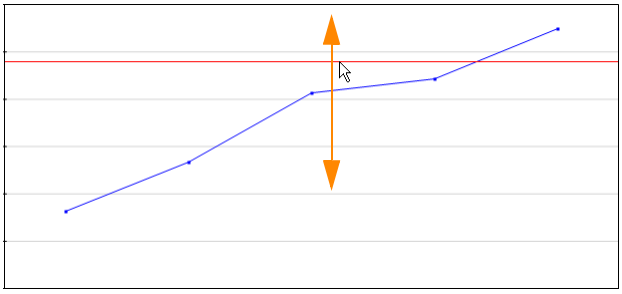
The Sample Thresholding window opens.

2 From the Columns list in the upper left of the window, select the first criteria you want to use to filter your data.

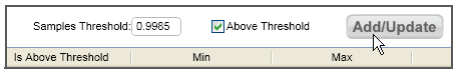
Available samples are shown as a line plot in the window.

3 Specify the minimum acceptable value for that criteria by:

- Dragging and dropping the red line up or down on the graph, or



- Typing a value in the **Threshold** text box below the graph.



Any samples above the threshold value (or the red line) will be included in your project or report automatically. Any samples below the threshold will be excluded.

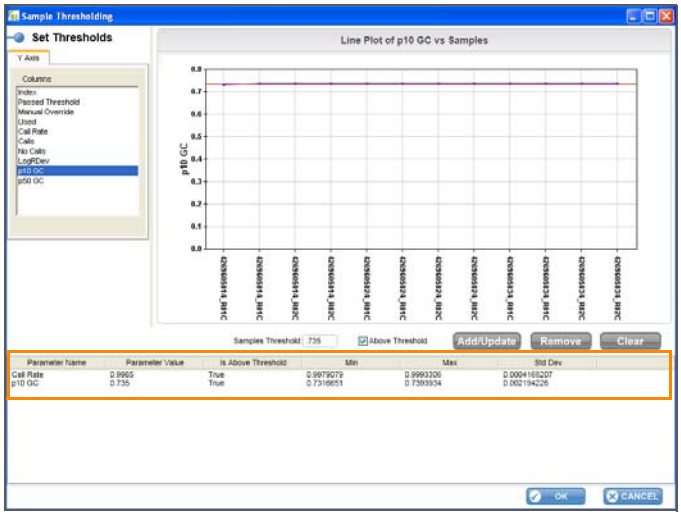
4 Click **Add/Update**.

The criteria you defined is summarized in a list in the lower portion of the window.

5 Repeat steps 2 through 4 to specify additional parameters to filter the samples in your data set.

In the example here, samples with a call rate below 0.9985 and a p10 GC value below 0.735 will be filtered out of your data set.

Figure 5 Call Rate and p10 GC Sample Threshold Parameters Set



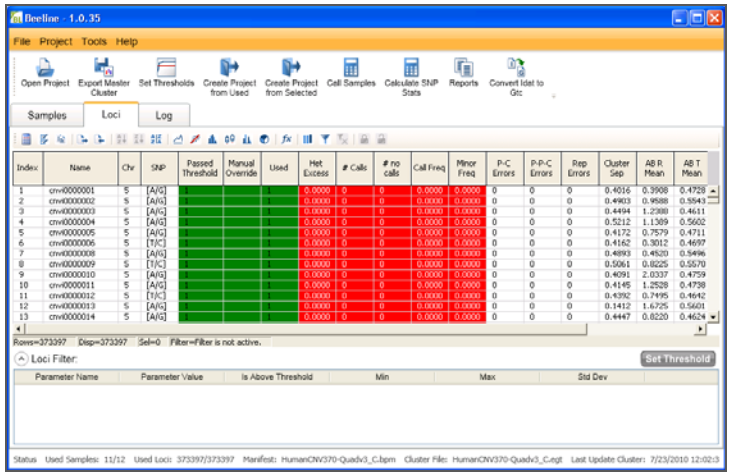
6 When you are finished, click **OK**.

Beeline reviews the samples in your data set against the criteria you set, then returns you to the Available Samples tab on the main screen.

The Passed Filter and Used columns for each sample are updated to indicate whether or not that sample will be included in a report or GenomeStudio project. If a sample passed filter, it shows a one (**1**) in both columns; otherwise, it shows a zero (**0**).

NOTE

If this project contains loci statistics, it is likely that the SNP statistics shown in the table on the Loci tab will be out of sync after filtering samples. When this occurs, the relevant columns on the Loci tab are highlighted red, as shown below. To update the table statistics to reflect "Used" data, click **Calculate SNP Stats**.



7 If you are happy with the results of the filter, you can:

- Filter loci,
- Generate a report, or
- Create a GenomeStudio project.

If you are not happy with the results of the filter, you can either:

- Adjust the threshold settings as described above, or
- Override the Used setting to include or exclude specific samples manually.

Manually Including or Excluding Samples

1 Select one or more rows in the Available Samples table. To select multiple rows, hold down the Ctrl key as you click each row.

2 With the row(s) selected, right-click and select **Manual override out** or **Manual override in**, as appropriate.

The Manual Override and Used columns are updated with ones (**1**) and zeroes (**0**) to indicate whether or not that sample will be included.

If you have thresholds set for automatic filtering, the value in the Used column will no longer match the value in the Passed Filter column.

3 Repeat steps 1 and 2 until the values in the Used column accurately reflect the samples you want to include in a report or GenomeStudio project.

NOTE

If this project contains loci statistics, it is likely that the SNP statistics shown in the table on the Loci tab will be out of sync after filtering samples. When this occurs, the relevant columns are highlighted red. To update the table statistics to reflect "Used" data, click **Calculate SNP Stats**.

Filtering Loci

Filtering loci is very similar to filtering samples. You can set thresholds to automate the filtering process or you can manually include or exclude specific loci.

Setting Thresholds to Filter Loci Automatically

The instructions below provide an example of how to filter loci based on two threshold criteria: call frequency and A BR mean. However, you can apply these instructions to filter on any criteria appropriate to your experiment.

1 With the Loci Statistics tab active, click **Set Thresholds**.

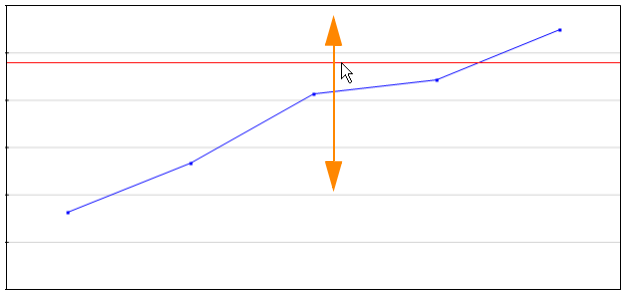
The Loci Thresholding window opens.

2 From the Columns list in the upper left of the window, select the first criteria you want to use to filter your data.

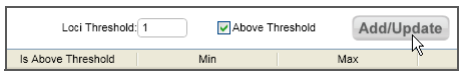
Loci in your data set are shown as a line plot in the window.

3 Specify the minimum acceptable value for that criteria by:

- Dragging and dropping the red line up or down on the graph, or



- Typing a value in the **Threshold** text box below the graph.



Any loci above the threshold value (or the red line) will be included in your project or report automatically. Any loci below the threshold will be excluded.

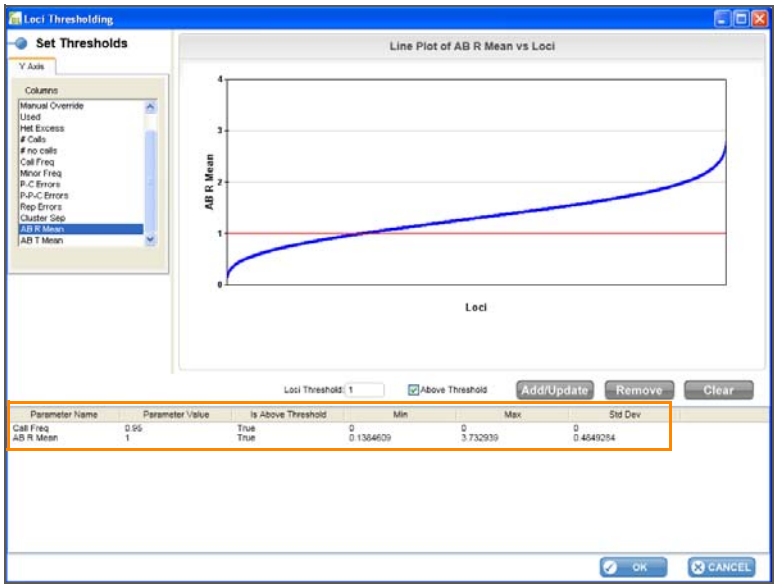
4 Click **Add/Update**.

The criteria you defined is summarized in a list in the lower portion of the window.

5 Repeat steps 2 through 4 to specify additional parameters to filter the loci in your data set.

In the example here, loci with a call frequency below 0.95 and an AB R mean value below 1 will be filtered out of your data set.

Figure 6 CallFreq and AB R Mean Loci Threshold Parameters Set



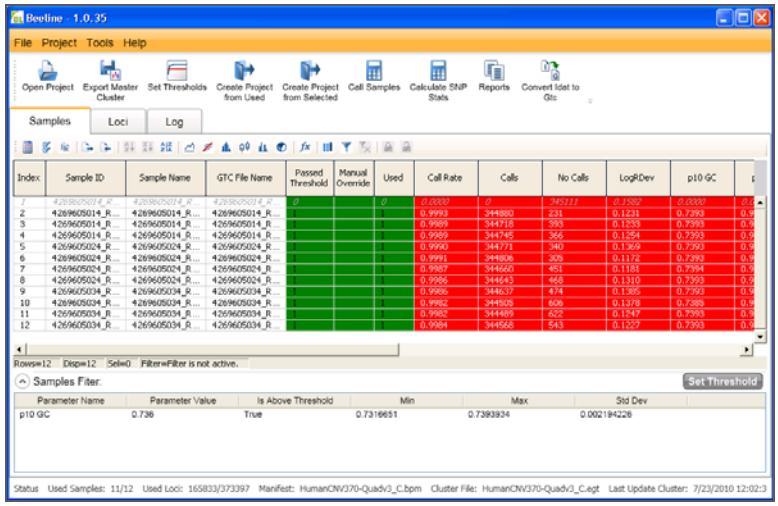
6 When you are finished, click **OK**.

Beeline reviews the loci in your data set against the criteria you set, then returns you to the Loci Statistics tab on the main screen.

The Passed Filter and Used columns for each locus are updated to indicate whether or not that locus will be included in a report or GenomeStudio project. If a locus passed filter, it shows a one (**1**) in both columns; otherwise, it shows a zero (**0**).

NOTE

It is likely that the sample statistics shown in the table on the Samples tab will be out of sync after filtering loci. When this occurs, the relevant columns on the Samples tab are highlighted red, as shown below. To update the table statistics to reflect "Used" data, click **Call Samples**.



7 If you are happy with the results of the filter, you can:

- Generate a report, or
- Create a GenomeStudio project.

If you are not happy with the results of the filter, you can either:

- Adjust the threshold settings as described above, or
- Manually override the Used setting to include or exclude specific loci.

Manually Including or Excluding Loci

1 Select one or more rows in the Loci Statistics table. To select multiple rows, hold down the Ctrl key as you click each row.

2 With the row(s) selected, right-click and select **Manual override out** or **Manual override in**, as appropriate.

The Manual Override and Used columns are updated with ones (**1**) and zeroes (**0**) to indicate whether or not that sample will be included.

If you have thresholds set for automatic filtering, the value in the Used column will no longer match the value in the Passed Filter column.

3 Repeat steps 1 and 2 until the values in the Used column accurately reflect the loci you want to include in a report or GenomeStudio project.

NOTE

It is likely that the sample statistics shown in the table on the Samples tab will be out of sync after filtering loci. When this occurs, the relevant columns on the Samples tab are highlighted red. To update the table statistics to reflect "Used" data, click **Call Samples**.

Modifying the Cluster File

Beeline does not enable you to modify cluster files directly. However, you can create a GenomeStudio project containing just your filtered data, modify the cluster file for that project in GenomeStudio, then merge the changes back into your Beeline project.

Creating a GenomeStudio Subproject

Beeline enables you to create a GenomeStudio project from a subset of your data based on threshold filters you set or based on manual selections in the Available Samples and Loci Statistics tables.

1 Do one of the following:
   - If the Used columns in the Available Samples and Loci Statistics tabs accurately reflect the data you want to include in the GenomeStudio project, click **Create Project from Used**.
   - Otherwise, highlight specific samples and loci on each tab to include in your project, then click **Create Project from Selected**.

   The Save Project dialog box opens. By default, Beeline will save the project in the GenomeStudio Working Folder you specified when you imported your data.

2 If you want to save the project in a different folder than the default, click **Browse** and navigate to the desired folder.

3 Specify a name for the project and click **OK**.

   Beeline saves the subset of data as a *.blc file and names the cluster file for this project subClusterFile.egt. The project opens in GenomeStudio.

Modifying Clusters in GenomeStudio

For information on modifying clusters in GenomeStudio, refer to the *GenomeStudio Genotyping Module User Guide*.

Merging Cluster Files

After modifying clusters in GenomeStudio, you must merge the changes back into the data set you have open in Beeline. You can merge the changes from within GenomeStudio or Beeline.

Merging Cluster File Changes in GenomeStudio

1 In GenomeStudio, select **File | Export Cluster Positions | For All SNPs**.

   GenomeStudio opens to the folder containing the subClusterFile.egt file for this project.

2 Select the subClusterFile.egt file and click **Save**.

3 When prompted to overwrite the cluster file, click **Yes**.

   The master cluster file open in Beeline will be updated automatically with the changes you made in GenomeStudio.

4 Return to the Beeline main screen.

   Your data set should still be open.

5 Select **Calculate SNP Stats** before filtering based on loci or generating reports.

Merging Cluster File Changes in Beeline

1. In GenomeStudio, select **File** | **Export Cluster Positions** | **For All SNPs**.
   GenomeStudio opens to the folder containing the subClusterFile.egt file for this project.
2. Type a new name for the cluster file and click **Save**.
3. Return to the Beeline main screen.
   Your data set should still be open.
4. Select **Project** | **Merge Cluster Subset**.
   The Select a Cluster File dialog box opens.
5. Navigate to the modified cluster file you created in GenomeStudio and click **Open**.
6. Select **Calculate SNP Stats** before filtering based on loci or generating reports.

Generating Reports

If you do not need to modify clusters or visualize your data in GenomeStudio, you can create reports directly in the Beeline software.

1 Click **Reports**.
The Report dialog box opens.

2 Select one or more of the following reports:
- **DNA Report**—Creates a snapshot of the samples currently showing in the Available Samples table.
- **Summary Report**—Provides a high-level overview of the project, including number of samples used, number of loci, and date.
- **Locus Summary Report**—Creates a snapshot of the loci currently showing in the Loci Statistics table.
- **Locus x DNA Report**—Provides calls and DNA sample information in a single report.
- **Final Report**—Provides allele calls of each locus for all samples.
- **Reproducibility and Heritability Error Report**—Provides an overview of error output.

3 Specify the number of samples you want to appear per file.

4 Click **Browse** and navigate to the folder where you want to save your reports.

NOTE
The Reproducibility and Heritability Error Report is saved in the GenomeStudio project folder, rather than the folder you specify here.

5 Specify a name for the collection of reports and click **OK**.
Each report is given a unique file name that starts with the name you specify, followed by an underscore and the type of report.
For example, if you specified the name **BC012345** and selected to generate a Locus Summary Report and a DNA Report, the reports would be named BC012345_Locus.csv and BC012345_Sample.csv.

Additional References

This section contains an overview of the system requirements for installing and running the Beeline Software, as well as pointers to supporting documentation and related reading materials that might be of interest.

System Requirements

To install and use the Beeline software, your computer must meet the following requirements.

CPU Speed	Intel Celeron Duo or faster
Processor	64-bit
Memory Size	8 GB or more
Hard Drive	100 GB or larger
Video Display	1280 x 1024
Operating System	Windows XP SP2, Vista, or 7
Specific OS Requirements	Microsoft .NET Framework 3.5
Network Connection	1 GbE or faster
Other Required Software	• iScan Control Software 3.2 or later • GenomeStudio 2010.2 or later

Supporting Documentation

You can download PDFs of the following documents from the Illumina website using your iCom login. Go to www.illumina.com/documentation.

- *GenomeStudio Framework User Guide* (part # 11318815)
- *GenomeStudio Genotyping Module User Guide* (part # 11319113)
- *iScan System User Guide* (part #11313539)

When you click on a document name, you will be asked to log in to iCom. After you log in, you can view or save the PDF.

Recommended Reading

The following reading materials are available on the Illumina website:

- "Improved Cluster Generation with GenTrain2" Technical Note
  www.illumina.com/Documents/products/technotes/technote_gentrain2.pdf
- Beeline Software web page
  www.illumina.com/software/beeline_software.ilmn
- GenomeStudio Software web page
  www.illumina.com/software/genomestudio_software.ilmn
- Beeline Online Training

Technical Assistance

For technical assistance, contact Illumina Customer Support.

Table 2 Illumina General Contact Information

Illumina Website	http://www.illumina.com
Email	techsupport@illumina.com

Table 3 Illumina Customer Support Telephone Numbers

Region	Contact Number
North America toll-free	1.800.809.ILMN (1.800.809.4566)
United Kingdom toll-free	0800.917.0041
Germany toll-free	0800.180.8994
Netherlands toll-free	0800.0223859
France toll-free	0800.911850
Other European time zones	+44.1799.534000
Other regions and locations	1.858.202.ILMN (1.858.202.4566)

MSDSs

Material safety data sheets (MSDSs) are available on the Illumina website at http://www.illumina.com/msds.

Product Documentation

If you require additional product documentation, you can obtain PDFs from the Illumina website. Go to http://www.illumina.com/support/documentation.ilmn. When you click on a link, you will be asked to log in to iCom. After you log in, you can view or save the PDF. To register for an iCom account, please visit https://icom.illumina.com/Account/Register.

Illumina, Inc.
9885 Towne Centre Drive
San Diego, CA 92121-1975
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America) techsupport@illumina.com
www.illumina.com